

~~UNCLASSIFIED-INTERNAL USE ONLY~~

IX. PROFESSIONAL APPLICANT TEST BATTERY (PATB)

The PATB, an eight hour series of written tests, has been used by the Agency for more than 20 years as a tool for the selection of new professional employees. It is administered by the Office of Medical Services/Psychological Services Staff (PSS). This inspection examined two aspects of the PATB -- its use in applicant processing and selection, and the evidence concerning its reliability and validity.

The first aspect was addressed by the inspection team, the second by two experts in psychological testing who have extensive experience in academic, military and industrial settings, and are widely quoted in current professional literature on the subject.

[The complete report prepared by the consultants appears as Appendix H to this report. Detailed consultant findings and explanations are necessarily omitted from this chapter which attempts to present their major findings in non-technical terms. Appendix H should be read in its entirety to obtain a full understanding of the consultants' findings.]

~~UNCLASSIFIED-INTERNAL USE ONLY~~

A. Agency Use of the PATB

Our findings concerning the use of PATB are based on a questionnaire survey of 900 employees entering on duty between 1 October 1977 and August 1979, and of 500 supervisors. They also derive from interviews of Agency employees, and from a review of Recruitment Guides which indicate whether the PATB is supposed to be administered to applicants for a particular type of job.

We find that there is no consistent policy within the Agency as to which applicants should take the PATB and which not. Among recently-hired professional employees, two thirds report they had taken the PATB and one third report they had not. All Directorates resort to it to some extent, but the Directorate of Science and Technology relatively little. It is, for example, administered to applicants for engineering positions in some components, but not in others. As applicant files are shunted from one component to another, it often happens that among applicants being considered for the same type of position, some will have been tested, some not.

ADMINISTRATIVE INTERNAL USE ONLY

Decision as to whether an applicant must take the PATB is made at the Office and Division levels, and such decisions often are not consistent among Offices and Divisions. Even within components which specify in their Recruitment Guides that the PATB is to be administered to applicants for certain positions, reliance on it by individual managers varies from extensive to total disregard. More than 60% of 500 supervisors who were surveyed either have no opinion about PATB's usefulness or indicate that it is not used by their components. Only about one-fourth of the supervisors indicate that they give significant weight to the applicant's performance on PATB in making employment decisions.

There is a need for the Agency to develop a systematic policy on the role of PATB in personnel selection. This does not imply that all applicants should be tested, just that applicants for a given type of position should be evaluated against uniform selection criteria. Policy guidelines need to be set for determining the positions for which the PATB is to be used as a selection device, and such guidelines should be followed by all components. If the decision to require or not require PATB for professional jobs is made arbitrarily by individual managers there is a high potential

ADMINISTRATIVE<sup>145</sup> INTERNAL USE ONLY

for violating EEOC Guidelines on disparate treatment of applicants. In the opinion of the Office of General Counsel, "If CIA does not comply with lawful and appropriate EEOC regulatory issuances, CIA would stand in violation of both statute and executive order<sup>1/</sup>."

RECOMMENDATION:

36. The Deputy Director of Central Intelligence instruct the Executive Committee to develop, for regulatory issuance, an Agency policy which specifies the types of positions for which selection tests are to be administered, and the types of testing appropriate to such positions.

B. Reliability

An important consideration in evaluating a test battery is its reliability; a test is regarded as reliable if repeated measurement gives consistent results for a

---

<sup>1/</sup> Memorandum from the Deputy General Counsel to the Director of Equal Employment Opportunity, OGC 79-05429, 13 June 1979.

REF ID: A66000  
ADMINISTRATIVE - INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130012-5

given individual or if the individual's relative standing in a group shows little change. If test scores are not reliable, they cannot be used with confidence for accurate measurement or prediction.

Some reliability data are available from PATB research performed in the mid-1950's. PSS states that in 1975, five additional studies were made of the reliability of the PATB; although only one 1975 study was made available to the consultants and the inspectors.

A review of these data shows that for white males and females, only five of the PATB's 31 tests have reliabilities that the consultants regard as being at least minimally acceptable. For all other tests and scales in PATB for which reliability data are available, they regard the reliabilities as below minimally acceptable level; i.e., the scores from them are too unstable for use in making decisions about individuals. No reliability data are available for the sample of writing ability or the scoring procedure. The consultants find PATB reliability data particularly inadequate for females and minorities, due apparently to failure to recognize that females and minorities have become a more significant portion of the Agency's professional work force than they had been in the 1950s.

The consultants caution that, until adequate evidence exists that a test is sufficiently reliable to be used, it is unwise as well as legally questionable to use it as a personnel selection device. To do so violates professional standards for test development and violates Equal Employment Opportunity Commission (EEOC) Guidelines because it can lead to false interpretation of test scores.

The Psychological Services Staff (PSS), in response to the consultants' findings and conclusions, asserts that when the PATB was developed in the 1950's it was validated, normed for use in personnel selection, and a concerted effort made to establish its reliability. PSS further asserts that once the reliability of the tests is established, and the test battery put into use, there is usually little reason to question the reliability further. We believe the changing composition of the applicant population constitutes ample reason to recheck PATB reliability. In consequence of the consultants' criticism of the adequacy of the reliability data, PSS recently conducted reliability studies of eight of the PATB's cognitive tests administered to 426 applicants in 1978 and 1979. On the basis of these studies, PSS advises that reliability co-efficients for the 235 males, 191 females, 228 whites and 198 blacks in the

study exceed minimally acceptable levels. We find several difficulties with these results: the statistical method used to establish test reliability yields spuriously high results when used with tests whose questions are not independent or for tests that are <sup>+</sup>timed. Also, results of reliability studies for the eight tests cannot be assumed to extend to the other 23 tests in the battery. A recommendation later in this chapter addresses the PATB's reliability within the context of a comprehensive approach to the Agency's psychological testing program. ✓

### C. Validity

A second major consideration in evaluating a test battery is its validity; a test is regarded as valid if there exists a demonstrably logical relationship between specific elements of the test battery and knowledge and/or skills required in the job, or there is a statistical relationship between test scores and performance on the job.

The consultants, in their review, considered the several types of validity recognized by the American Psychological Association (APA). Essentially, there are three -- content validity in which actual elements of the

work are replicated in testing; construct validity in which knowledge, abilities and skills logically related to tasks performed on the job are tested; and criterion-related validity in which statistical comparisons are made between individuals' scores on certain tests (not necessarily related demonstrably to the work) and measures of incumbents' performances on the job. An example would be a correlation between scores on a numerical operations test and performance in a job not involving mathematical computation. Virtually all of the validity studies conducted by PSS on elements of the PATB over the years have been of the criterion-related type.

The consultants reviewed 23 studies done by, or under the auspices of, PSS. They conclude that the evidence presented for the validity of these studies is seriously inadequate. They state that of 16 PSS studies relating performance on the tests to performance on the job, only ten provide sufficient information to permit judging the evidence for validity. They find that even the evidence which is presented in the ten studies is fragmentary, very weak and unconvincing, and does not meet minimum standards set by the APA or the EEOC's Uniform Guidelines on Employee Selection Procedures. The consultants point out that the samples



of job incumbents which were used are too small; that the incumbents who were studied represent primarily a high ability range and cannot effectively be used to validate tests which seek to differentiate among applicants across the entire range of abilities; and that relevant, reliable and unbiased job performance measures were not available by which to judge the incumbents' success on the job.

The consultants also find that PSS' statistical measures of the degree of agreement between the test scores of individuals who were hired, and supervisors' subsequent rating of the performances of these individuals in particular jobs have been inconsistent and generally low. They assert that such results are to be expected in the absence of suitably reliable tests and when the hiring process eliminates the less talented applicants, resulting in a restriction in the range of talent available for subsequent study.

PSS has, in some of its studies, developed special criteria for measuring employee performance, but PSS acknowledges that the job performance data used in its validation studies generally is not satisfactory. Fitness report ratings on which many of the studies are based are neither ✓

sufficiently reliable nor differentiated to provide an acceptable basis for test validation.

The consultants find no validity data of any kind for the PATB writing sample and point out that the Strong-Campbell Interest Inventory (SCII), which compares an applicant's self-expressed interests with those of incumbents in a wide variety of jobs common outside the Agency, lacks both Agency norms and Agency validation. PSS contends that Agency norms per se are not necessary for the SCII because external interest profiles developed as part of the instrument are relevant to Agency jobs. The inspectors find, however, that a hiring official contemplating the selection of an applicant for a job as a librarian, researcher, or security investigator is not demonstrably helped when informed in the PATB narrative report that the applicant has interests similar to a forester or aviator. ✓ Although there have been some attempts to validate certain items on the Biographical Information Inventory, statistical evidence for its validity is lacking and, again, because of the lack of explicit job analyses, logical relationships cannot be established.

Four PSS studies have been done on the validity of PATB test scores for predicting success in foreign language

training. According to the consultants, only two of the 31 PATB tests were related with any consistency to success in foreign language training but these related only to training in French and Spanish. They conclude that the results of these four studies do not establish the value of PATB for evaluating language aptitude. PSS objects to this conclusion and states its work in this area remains exploratory. Still we find the results being reported to hiring officials as predictors of performance in foreign language study.

It is the consultants' view that the confidence with which results of the validation studies are reported as predictors of job performance greatly exceeds the level of confidence which is justified by the statistical data on the reliability and validity of the PATB.

Given the inadequacies, on both theoretical and practical terms, of the existing criterion-related test validation program, we believe the Agency should adopt a distinctly different approach which is both professionally and pragmatically sound. The consultants recommend a construct test validation based on comprehensive job analyses, and we agree. (A proper job analysis identifies

ADMINISTRATIVE USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130012-5

the knowledge, abilities and skills an individual should possess in order to perform a particular job, or group of similar jobs, effectively.)

The consultants find no evidence, in the PSS data which were made available to them, that comprehensive job analyses have been performed for the purpose of validating the PATB or other Agency personnel selection procedures. They explain that the mean test score profiles of individuals already performing Agency jobs, which are now relied upon, do not constitute job analyses. They conclude that there is no logical, professionally justifiable relationship (or construct validity) between the PATB and the jobs for which it is used as a selection tool. PSS asserts that its Test Data Book #15, dated 1 July 1958, constitutes evidence of job analyses appropriate for this purpose. The consultants disagree, pointing out that almost all the data in the book relates to success in training rather than to success on the job and that <sup>training</sup> criteria cannot be used to demonstrate job-related validity. ✓

The consultants cite the absence of job analyses upon which to base selection as the most serious deficiency of the PATB. They assert that the lack of job analyses is a violation of APA professional test development standards

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130012-5

ADMINISTRATIVE USE ONLY

and of EEOC Guidelines when a statistical relationship between test scores and job performance ratings cannot be established feasibly.

PSS, in commenting on the consultants' report, has construed their recommendation of comprehensive job analyses for the Agency as a "mechanistic, task-oriented approach". The thrust of the consultants' recommendation is that comprehensive job analyses identify not the detailed tasks to be performed, but the human attributes (knowledge, abilities, and skills) needed to perform a given job, or group of similar jobs. In endorsing the consultants' view, the inspection team is suggesting the use of just such a job analysis instrument, one which has been used widely, although certainly not exclusively, in private industry and which forms a basis for the Federal Bureau of Investigation's testing program to select new special agents.

Initial work in conducting comprehensive job analyses could be accomplished under contract by a group of cleared test development specialists/consultants, using commercially available instruments. At the same time, the Agency should hire test development specialists for an ongoing capability. We estimate that three to four professional personnel, with supporting staff, would be a workable complement for such a

test development unit. Appendix H of this report should be provided to them for initial guidance.

The inspection team believes that responsibility for the development of an Agency comprehensive job analysis program should be placed within the Office of Personnel Policy, Planning, and Management, rather than in the Psychological Services Staff, Office of Medical Services, because such analyses are relevant not only to testing but to the entire range of personnel selection procedures. We note that a recommendation to this effect recently was made by the Agency task force for implementing the Uniform Guidelines and was approved by the DDCI.

RECOMMENDATIONS:

37. The Deputy Director of Central Intelligence:
  - a. authorize the Director of Personnel Policy, Planning, and Management to contract with job analysis specialists to assist the the Agency in developing an Agencywide job analysis program by a specific date. These specialists should consider the Position Analysis Questionnaire, developed at Purdue University, for this purpose. It is a worker-oriented, as opposed to a task-oriented,

approach. It permits one to generalize the human attributes underlying diverse jobs performed in the Agency.

- b. establish a unit of three to four professionals under the Director of Personnel Policy, Planning, and Management to develop and try out a new applicant testing program, and establish professionally acceptable reliability and validity data and norms for such tests before they are authorized for administration and use in personnel selection. This unit should also be responsible for assuring the reliability and validity of all other Agency testing for professional and nonprofessional applicants. Job analysts and test development specialists with demonstrated professional training and practical experience in the specialized fields of job analysis and test development should be hired for this unit.

D. PATB Narrative Report

The consultants have major reservations with respect to the narrative report that is prepared to summarize an applicant's PATB results. They assert that the strong, confident recommendations to hire or not to hire applicants for specific jobs or in specific components are not supported by the available evidence for validity of the PATB. The consultants point out that since most of the PATB tests lack adequate reliability and have little demonstrated validity for jobs in the Agency, the narrative reports based on the test scores are misleading and potentially unfair. In addition, they find no written guidelines available to or used by the psychologists in writing the narrative report, and state that sections of the reports tend to vary considerably in unpredictable ways. To them, the variations appear due as much or more to the personal idiosyncrasies of the psychologists as to differences in performance on PATB among applicants. For example, reports of applicants' writing abilities variably address grammar, syntax, spelling, sense, and literary quality, and use ambiguous terms to describe the results.



ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130012-5

PSS challenges criticism of the PATB narrative report because the consultants and inspectors failed to examine the raw test scores on which the reports are based. The Staff asserts that objective judgments of the reporting of test results cannot be made in the absence of direct, side-by-side comparison of the narrative reports with the test scores on which those reports are based. The inspection team acknowledges that such side-by-side comparison was not made but reiterates that comparable sections of narrative reports vary greatly in both the selection and treatment of points which are addressed. In the final analysis, the real value of the narrative report to a hiring official depends on the reliability and validity of the test results behind it, and on their objective presentation.

#### E. Relevance for Minorities and Females

At the time of the consultants' review there was no evidence that studies of adverse impact as defined by the EEOC had been done for PATB or for any other selection procedure used in the Agency. Although there was no direct evidence of bias or unfairness, the consultants believe there is the following serious potential for misuse or unfair use of PATB:

ADMINISTRATIVE-INTERNAL USE ONLY

**ADMINISTRATIVE INTERNAL USE ONLY**

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130012-5

- no evidence that minorities were represented in the groups used to norm the tests;
- no evidence that minorities have been included in the samples used to determine job-related validity;
- evidence that females have not been represented at all in some samples and are under-represented in others;
- reliability data for PATB tests and scales are not available for minorities, nor for the work attitude scales for females.

Only two studies of minority applicants apparently have been done by PSS. According to the consultants, one of these, done in 1974, did not analyze the data correctly and must be disregarded. The other, initiated in 1979, they view as not yet conclusive with respect to fairness of the PATB. The study was in process when they completed their review. PSS reports that the now completed study of 952 black applicants between January 1974 and January 1977 reveals that the Agency hired approximately the same percentage of blacks from among those who took the PATB as it did from among those who did not. The PSS study also found that, among black applicants who were tested, the

**ADMINISTRATIVE INTERNAL USE ONLY**

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130012-5

scores of those who were hired were superior overall to the scores of black applicants not hired. Finally, PSS reports that during this period the Agency hired 15% of the black applicants who were tested, 13.8% of the white applicants who were tested. PSS concludes that PATB testing has no adverse impact on black applicants and that this study is sufficient to meet EEOC Guidelines on this question.

We conclude from our own review of the Uniform Guidelines for Employee Selection Procedures that PSS' view on this question should be acceptable to the EEOC for the time being. However, the Guidelines stipulate that data about the impact of selection techniques, including tests, must now be compiled in relation to specific jobs or job categories. This is not yet being done.

PSS asserts that there is not enough test data for other minority groups to make any determination whether these groups are experiencing adverse impact from the PATB. Again, the Guidelines specify that in the absence of such data, "the Federal enforcement agencies may draw an inference of adverse impact of the selection process from the failure of the user to maintain such data, if the user has an underutilization of a (minority) group in the job category, as compared to the group's representation in

the relevant labor market or, in the case of jobs filled from within, the applicable work force." (Section 4.D.) Thus, the Agency will be required to maintain and report minority test data with which user components will be able to conform to the Guidelines.

F. Summing Up

The Psychological Services Staff seeks an augmentation of its staff with which to intensify its research work on the PATB and other psychological services; it also seeks improved access to personnel data which would strengthen its criterion-related test validation program. We believe that PSS is, indeed, shorthanded, but are concerned that, in its comments on the consultants' study, appears to support the status quo concerning the content, reliability and validity of the PATB as well as to endorse individual managers' completely discretionary use of it in the selection process.

By contrast, we believe that major changes should be made in the testing program to make it more reliable and to convert test validation to the construct validity concept based on a comprehensive job analysis program within the Agency. We are moved to this view both by the consultants'

report and by the Uniform Guidelines which call for selection criteria based on job analysis.

Despite their criticism of the PATB and of the validity evidence for it, the consultants see the Agency's need for a battery of good selection tests. If testing is eliminated, the only procedures for selecting personnel would be interview, review of past academic records and experience, and personal recommendation. They view each of these alternative procedures as having major shortcomings, with none being able to provide the sort of relevant information about an applicant's capabilities that are potentially available from a good selection battery.

Based on the consultants' foregoing analysis of PATB and their evaluation of the Agency's vulnerability to legal challenge, we considered the recommendation that PATB be suspended in its entirety until validity and adequate reliability of its tests have been established. However, we share the consultants' belief that the Agency needs a battery of good selection tests and recognize the possibility that future studies may indeed confirm the construct validity and adequate reliability of some of the existing PATB tests. Consequently, we favor a modified course of action which retains certain aspects of the current PATB

program while efforts are initiated to develop a sounder psychological testing program for the Agency.

Although the consultants found no suitable evidence for validity of the measures of intellectual ability, and that some of these tests had low reliabilities, we do not recommend that use of these be discontinued at this time. Some require modification, and all require new norms based on a representative sample of current applicants. They also require a logical rationale (to establish construct validity) for their use, based upon a sound job analysis. Pending such developments, they may provide at least some basic measure of an individual's intellectual ability. They should be modified, however, with all possible speed.

RECOMMENDATION:

38. The Director of Medical Services continue to administer the following tests subject to the Agency's initiating a job analysis and test development program:

- Vocabulary
- Reading Comprehension
- Figure Matrices
- Arithmetic Reasoning

- Contemporary World Affairs
- Numerical Operations
- Considerations
- Interpretation of Data
- Essay (Writing Sample)

We agree with the consultants that profile results of the eight cognitive tests listed above should routinely be reported in the files of all applicants who are tested, and that names of applicants with special skills and high abilities be retained for computerized recall for an indefinite period. Reports of applicant test profiles should include the notation that the use of the eight cognitive tests is an interim procedure pending a validated testing program. The essay also should be included in the applicant file on an unevaluated basis. This data, of course, should be removed when an applicant enters on duty.

RECOMMENDATIONS:

39. The Director of Medical Services, in consultation with the Director of Personnel Policy, Planning, and Management, insure that a profile of cognitive test results, plus the unevaluated essay, be placed in an applicant's file prior to its review by hiring officials.

40. The Director of Personnel Policy, Planning and Management enter into computerized records for indefinite retention the names, special skills and test profiles of hard-to-get and unusually promising applicants.
41. The Director of Medical Services discontinue reporting test results for the following PATB tests and procedures which the consultants view as indefensible:
- Strong-Campbell Interest Inventory
  - Language Aptitude
  - Work Attitudes
  - Thurstone Temperament Schedule
  - Biographical Inventory
  - PSS Professional Applicant Testing Report  
(the narrative report written by OMS/PSS psychologists) in its entirety.

The tests could, however, continue to be administered for internal research purposes until sufficient validity data are available to support their use as a selection tool.